
Dependence Structure Estimation via Copula

Jian Ma* Zengqi Sun

Department of Computer Science and Technology
Tsinghua University
Beijing 100084, China
{majian03@mails,szq-dcs@mail}.tsinghua.edu.cn

Abstract

We propose a new framework for dependence structure learning via copula. Copula is a statistical theory on dependence and measurement of association. Graphical models are considered as a type of special case of copula families, named product copula. In this paper, a nonparametric algorithm for copula estimation is presented. Then a Chow-Liu like method based on dependence measure via copula is proposed to estimate maximum spanning product copula with only bivariate dependence relations. The advantage of the framework is that learning with empirical copula focuses only on dependence relations among random variables, without knowing the properties of individual variables. Another advantage is that copula is a universal model of dependence and therefore the framework based on it can be generalized to deal with a wide range of complex dependence relations. Experiments on both simulated data and real application data show the effectiveness of the proposed method.

1 Introduction

Dependence between random variables is of fundamental importance because it may imply essential statistical relations within real world social, physical, or biological systems. A large amount of data sets are collected from different fields, such as biology, social networks, finance, world-wide web. Analysis on them remains a challenge. Hence, dependence structure learning is one of the most contributed problems in the machine learning community.

The most well established statistical methodology for

dependence representation is graphical models, or Bayesian networks (Heckerman et al., 1995; Buntine, 1996; Jordan, 1998). Through graphical models formalism, a probability density is represented with a directed or undirected graph, in which each node represents a random variable, and each edge represents a conditional dependence relation between two random variables. Therefore, representational simplicity, through bivariate dependence decomposition, reduce computational complexity and makes large-scale problem modeling and inferring tractable. The implicit assumption of Graphical models is markovity or conditional independence, which means only first order dependence or pairwise dependence is considered in models. But the approach may be improper in many cases.

On the other side, traditional methods on inferring graphical models always involve maximum likelihood, where we should specify parametric family of entire underlying distribution, including margins of individual variables implicitly. Hypothesis selection on margins is central to the performance of structure learning to a large extent. But there is short of priori knowledge needed for such selection. So we are interested in finding a method in that can separate structure learning from parametric marginal specification.

Copula theory unifies the representation of multivariate dependence (Joe, 1997; Nelsen, 1998). The term “copula”, come from latin, refers to the way that random variables relate to each other. According to Sklar theorem (Sklar, 1959), multivariate distribution can be represented as a product of its margins and a copula function which represents dependence structure among random variables. Using copula, one can separate the margins from their joint density distribution, and therefore study only statistical interrelations without knowing individual properties of each variable. Copula has a wide applications in finance (E. Bouyé & Roncalli, 2000), and recently gain the notice of machine learning community (Ma & Sun, 2007; Kirshner,

Footnote for author to give an alternate address.

2007).

The main contribution of the paper is introducing a novel framework of structure learning based on copula. We study estimating dependency structure via copula. Particularly, we propose that dependence structure is first approximated by empirical copula (or copula density) and then fit certain dependence model on it. The most advantage of empirical copula is that it is a rank-based and model-free non-parametric estimation of underlying ‘true’ copula. Based on empirical copula estimation, many dependence structures can be further adopted and approximately inferred. For instance, in this paper, graphical models can be identified as a special case of copula. Graphical models concerns only pairwise dependence, and has its counterpart in copula theory, called product copula. We propose inferring product copula by Chow-Liu like algorithm (Chow & Liu, 1968) based on empirical copula estimation. Moreover, our methods can be generalized to estimate much flexible relationships.

2 Copula and Copula Space

2.1 Definition and Properties

Copulas are the functions that model the dependence relations among random variables, and can be defined as follows:

Definition 2.1 (Copula). (Joe, 1997; Nelsen, 1998) Given N random variables $\mathbf{X} = \{X_1, \dots, X_N\} \in \mathcal{R}^N$. Let $\{u_i = F_i(x_i), i = 1, \dots, N\}$ be the marginal distributions of \mathbf{X} . A N -dimensional copula $C : \mathcal{I}^N \rightarrow \mathcal{I}$ ($\mathcal{I} = [0, 1]$) of \mathbf{X} is a function with following properties:

- C is grounded and N -increasing;
- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$.

Intuitively, copula can be viewed as a new cumulative distribution function (CDF) stretched onto $\mathbf{u} = \mathcal{I}^N$ from the CDF of \mathbf{X} .

The relation between CDF, margins, and copula is stated in Sklar’s theorem (Sklar, 1959):

Theorem 2.2 (Sklar’s Theorem). *Given a random vector $\mathbf{X} = \{X_1, \dots, X_N\}$, its CDF $\mathbf{F}(\mathbf{x})$ can be represented as*

$$\mathbf{F}(\mathbf{x}) = C(u_1, \dots, u_N), \quad (1)$$

where C is a copula function, $\{u_i\}$ are marginal distribution functions of \mathbf{X} . If $\{F_i\}$ are continuous, then C is unique.

Sklar’s theorem is of fundamental importance in copula theory. By applying derivative on equation (1), we

can also represent probability density function (PDF) via copula. But Let us first present a new definition named *copula density*.

Remark 2.3. According to Sklar’s theorem, dependence structure is dependent from margins. This implies that it is possible that a same structure is learned from different distributions. These distributions are said to be *equivalent* in a sense of copula.

Definition 2.4 (Copula Density). A N dimensional copula density \mathbf{c} corresponding to N -copula \mathbf{C} is defined as

$$c(\mathbf{u}) = \frac{d^N}{du_1, \dots, du_N} C(\mathbf{u}), \quad (2)$$

where $\mathbf{u} \in \mathcal{I}^N$.

With the definition of copula density, we can derive a corollary of Sklar’s theorem:

Corollary 2.5. *The probability density function (PDF) $p(\mathbf{x})$ of \mathbf{X} can be represented as:*

$$p(\mathbf{x}) = c(\mathbf{u}) \prod_{i=1}^N p_i(x_i) \quad (3)$$

where $\{p_i, i = 1, \dots, N\}$ are marginal density functions of \mathbf{X} , and c is copula density.

2.2 Copula Space

As dependence structure representation, copula functions compose of a convex set enclosed by Minimal copula and Maximum copula (Nelsen, 1998).

How to construct a multivariate copula is of importance in applications. Despite Sklar’s theorem guarantees the existence of a copula function, it can not always be identified as a parametric one. In many cases we cannot write down an analytic copula. The following results provides the ways of constructing flexible copula representations for multivariate cases.

2.2.1 Mixture of Copulas

Theorem 2.6 (Mixture of Copulas). *The geometric mean of copulas (or copula densities) is also a copula (or copula density).*

The theorem can be illustrated as follows:

$$c(\mathbf{u}) = \sum_{k=1}^K w_k c_k(\mathbf{u}), \quad (4)$$

where c_k represents any type of copula density, and $\sum_{i=1}^K w_i = 1, w_i \geq 0$.

Remark 2.7. Based on the above result, we can construct more flexible copula model by mixture of copulas. These copulas being mixed together can be from parametric families, or product copulas to be presented below (Kirshner, 2007).

2.2.2 Product Copula

Theorem 2.8 (Product copula). *The product of copula density of independent variables is also a copula density.*

The theorem can be illustrated as

$$c(\mathbf{u}) = \prod_{m=1}^M c_m(\mathbf{u}_m). \quad (5)$$

where $\{c_m\}$ are any type of copula density, and $\mathbf{u} = \cup_{m=1}^M \mathbf{u}_m$, and $\{\mathbf{u}_m\}$ are vectors of marginal functions of random variables. If all the sub-copulas c_m are bivariate, it means that there is only pairwise dependence exists. In this case, product copula is equal to a graphical model.

Theorem 2.9. *Any graphical models equals to a product copula with only bivariate sub-copulas.*

The theorem indicates that graphical model is just a special case of product copula. More generally, hypergraph can also be formulated to be a special case of product copula with each variable dimension sub-copula corresponding to a sub-graph.

3 Estimation Methods via Empirical Copula

A large mount of inference method for copula can be summarized as follows: starting with a parametric family of copula, either implicitly implied by pdf or explicitly specified, and then optimizing parameters under the maximum likelihood framework. Using nonparametric method will help us avoid the risk of parametric model family when no priori knowledge is available. In this section, we introduce empirical copula (density) estimation algorithm. To the best of our knowledge, there is no such works before.

3.1 Empirical Copula

Empirical copula was introduced by Deheuvels (Deheuvels, 1979; Deheuvels, 1981). It approximate the copula or copula density of samples based on order statistics.

Consider a i.i.d. sample set $X^t = \{x_1^t, \dots, x_N^t\} \in \mathcal{R}^N, t = [1, T]$. Let $\{x_n^{(t)}\}$ be order statistics and the corresponding rank $1 \leq r_n^t \leq T$ so that $x_{r_n^t}^t = x_n^{(t)}$.

Definition 3.1 (Empirical Copula). An Empirical copula $\hat{\mathbf{C}}$ of samples $\{X^t, t = 1, \dots, T\}$ is defined on a $(T+1)$ lattice

$$L = \left\{ \left(\frac{t_1}{T}, \dots, \frac{t_N}{T} \right) : t_n \in [0, \dots, T], n = 1, \dots, N \right\}, \quad (6)$$

as following:

$$\hat{\mathbf{C}} \left(\frac{t_1}{T}, \dots, \frac{t_N}{T} \right) = \frac{1}{T} \sum_{t=1}^T \mathbf{I}_{[r_n^t \leq t_n]}, n = 1, \dots, N. \quad (7)$$

where \mathbf{I} denotes indicator function.

Using forward difference on lattice, empirical copula density can be derived in a same way:

$$\hat{c} \left(\frac{t_1}{T}, \dots, \frac{t_N}{T} \right) = \sum_{i_1=1}^2 \dots \sum_{i_N=1}^2 (-1)^{\sum_n i_n} \hat{\mathbf{C}} \left(\left[\frac{t_n - i_n + 1}{T} \right] \right). \quad (8)$$

3.2 Estimation Algorithm

Based on the definition (7), it is not hard to present the estimation algorithm (see algorithm 1) of empirical copula, given a group of samples. According to equation (8), the algorithm of empirical copula density function is just a accumulative process based on algorithm 1.

The algorithm 1 has the linear time complexity $\mathbf{O}(TN)$ while the algorithm 2 $\mathbf{O}(TN * 2^N)$ a exponential one. Notice that both algorithms are based on samples. We can calculate empirical value of both functions as search table in advance, which can reduce the calculation time while applications.

Algorithm 1 Empirical Copula Function $\hat{\mathbf{C}}$

Input: data x_i , dimension N , size T , $\mathbf{u} \in I^N$
for $n = 1$ **to** N **do**
 $r_n = \text{Rank}(x_n)$
end for
 $m = 0, u_n = u_n * T$
for $t = 1$ **to** T **do**
 Initialize $n = 1$
 while $u_n \leq r_n^t$ **do**
 $n = n + 1$
 end while
 if $n = N + 1$ **then**
 $m = m + 1$
 end if
end for
Output: $\hat{\mathbf{C}}(\mathbf{u}) = m/T$

3.3 As a universe basement

Using empirical copula when estimating dependence structure have many advantages. First, with nonparametric empirical copula algorithm, we can estimate different dependence relations from data in a model-free way. Second, copulas are invariant under monotonically increasing transformation so we do not have to

Algorithm 2 Empirical Copula Density Function $\hat{\mathbf{c}}$

Input: data x_i , dimension N , size T , $\mathbf{u} \in I^N$
 $\hat{\mathbf{c}}(\mathbf{u}) = 0$
for all $\mathbf{t} \in [1, 2]^N$ **do**
 $\hat{\mathbf{c}}(\mathbf{u}) = \hat{\mathbf{c}}(\mathbf{u}) + \hat{\mathbf{C}}(\mathbf{u} - \frac{\mathbf{t}-1}{T})$
end for
Output: $\hat{\mathbf{c}}(\mathbf{u})$

normalize data during analysis. Third, copulas are insensitive to outliers.

4 Maximum spanning copula estimation by Chow-Liu algorithm

In this section, we want to go further step to include structure learning by graphical model into our copula framework. As previously stated, Graphical model is a special case of copula. The dependence relation represented by edges in graph is equal to a product of a group of bivariate copulas. In this section, we propose inferring such product copula from data by Chow-Liu algorithm based on empirical copula estimation. Notice that copula has all the dependence information of random variables. Inferring only product copula from data is just a way of approximating the ‘true’ underlying copula.

4.1 Maximum spanning copula problem

We want to estimate the dependency structure without being bothered by individual variables’ properties. Dependence is measured by copula, if possible, only by copula such that we don’t have to make additional assumptions on and inference the parametric form of individual variables under the risk of unfit models.

Suppose we want to approximate dependence relations with a type of structure $\mathbf{T}(\mathbf{t})$, where \mathbf{t} is the parameter specifying \mathbf{T} . Given a group of i.i.d. samples X generated from a N dimensional random vector $\mathbf{x} \in \mathcal{R}^N$ $p(x)$, an objective function F can be defined on it, and be optimized to inference \mathbf{T} with respect to \mathbf{t} :

$$\min_{\mathbf{t}} F(\mathbf{t}; X). \quad (9)$$

In many works, objective function F is defined through maximum likelihood principle, which requires parametric assumptions on multivariate density functions $p(x)$.

Now consider a N dimensional copula density \mathbf{c} of \mathbf{x} . We can derive its empirical estimation $\hat{\mathbf{c}}$ based on X , which contains all the dependence information in data. One of the natural idea is to cover the most dependence relations with product copula. We call such

product copula “*maximum spanning copula*” (MSC). Therefore, we transform structure learning into a fitting problem.

In this paper, we focus on MSC by product copula. The MSC approximation of \mathbf{c} in terms of bivariate product copula (or graphical model) composes of a product of $N - 1$ bivariate copula. Then the above F can be defined by the sum of dependence measurements on $N - 1$ subcopulas. The problem is how to find such a optimal product copula to approximate the dependence structure among random variables.

4.2 Dependence measure

First, we should choose a dependence measurement by copula.

4.2.1 Statistical measures of independence

Copula summarizes all the dependence relations. Hence it is natural to link it with the proposed dependence measures in statistics. It has been proved that measures, such as Kendall’s tau, Spearman’s rho, Gini’s gamma, can be calculated through only copula function (Nelsen, 1998). Using empirical copula (density) to approximate copula (density), we can calculate these measures approximately. For example, an estimation of Spearman’s rho based on empirical bivariate copula is

$$\rho = \frac{12}{T^2 - 1} \sum_{t_1=1}^T \sum_{t_2=1}^T \left[\hat{\mathbf{C}}\left(\frac{t_1}{T}, \frac{t_2}{T}\right) - \frac{t_1 t_2}{T^2} \right], \quad (10)$$

where T is the order of lattice.

4.2.2 Mutual Information

Mutual Information (MI) is dependence measure in information theory (Cover & Thomas, 1991). Due to copula density is actually a density on I^N , MI can be used to measure the divergence between the ‘true’ densities and its estimations.

$$I(x, y) = \int_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy. \quad (11)$$

The equation (11) can be transformed into a copula density representation:

$$I(x, y) = \int_{x, y} p(x)p(y)\mathbf{c}(u_x, u_y) \log \mathbf{c}(u_x, u_y) dx dy, \quad (12)$$

where u denotes marginal distributions. Given a group of data (x_1, x_2) , we can calculation MI as

$$I(x, y) = \sum_{(x_1, x_2)} \hat{p}(x_1)\hat{p}(x_2)\hat{\mathbf{c}}(u_x, u_y) \log \hat{\mathbf{c}}(u_x, u_y). \quad (13)$$

In this formulation, besides empirically estimated copula density, univariate marginal densities are the functions to be estimated, for which there are many well-established methods, such as naive estimator, k-NN, kernel methods, etc.. (Silverman, 1986). For this problem, we adopt gaussian kernel estimator due to its easily calculation of both density and its derivative.

4.3 Construction algorithm of product copula

We propose approximating from samples their copula (density) in form of product copula (density). First, based on dependence measure matrix, an complete graph \mathcal{G} on N random variables is built where the weight of each edge is equal to dependence between two variables. Constructing optimal product copula equals to finding maximum spanning tree $\}$. This is a well-defined problem, which can be solved by Chow-Liu algorithm (Chow & Liu, 1968). Chow-Liu algorithm is actually an algorithm for constructing maximum spanning tree with MI as edge weights. There are some established algorithms, such as Kruskal’s algorithm (Kruskal, 1956) and Prim’s algorithm (Prim, 1957). Both algorithms can find the solution in polynomial time. We adopt Prim’s algorithm in our method. It starts with an edge set E containing only the maximum weight edge, and then each time add from the complement set of E one vertex u and its corresponding edge (u, v) with maximum weight such that $v \notin E$ has edge connection with E and (u, v) will makes no loop in new E , till E contains all the vertex, in the case of complete graph also means $N - 1$ edges.

4.4 Algorithm

We give the whole algorithm in the section, which composes of three steps:

Algorithm 3 Estimating dependence structure via copula

Input: data x_i , dimension N , size T , $\mathbf{u} \in I^N$
Construct empirical copula density $\hat{\mathbf{c}}$ by algorithm 2;
Calculate MI matrix $\mathbf{M}_{\mathbf{x}}$ of \mathbf{x} by Equation (13);
Build dependence tree \mathbf{T} by Chow-Liu algorithm based on $\mathbf{M}_{\mathbf{x}}$.

4.5 Related to Density Estimation

If we want to estimate not merely dependence structure but the whole underlying density, nodes and edges in graph should be parameterized after graph is derived for dependence structure. Because copula is separated margins from distribution, density estimation

here can be achieved by estimation on marginal additional to structure learning. What we should do is to estimate one dimensional margins of individual variables. There are many established methods on this issue, which is however beyond the topic of this paper.

5 Experiments and Results

5.1 Simulated data

We perform our method on a group of simulate multivariate data to investigate the effectiveness of our method. A dataset with 1000 samples are randomly generated from a 5 dimensional distribution of a random vector, of which the first three elements are zero mean Gaussian and the others two are governed by Gaussian copula with margins as normal distribution and exponential distribution respectively. Due to only gaussian and gaussian copula exists, we measure dependence with Spearman’s rho using Equation (10). Based on estimated measurements, a graph is expected to be estimated where three gaussian variables and two variables coupled by copula are grouped together respectively.

The algorithm 3 is run on the dataset and then a empirical copula representing dependence relations between random variables are estimated which is illustrated in Figure 1. Based on it, we derived a approximate dependence tree as illustrated in figure 2.

5.2 Real datasets

The success of simulation experiments on toy data can be easily anticipated. We perform our method on two real datasets: abalone and housing from UCI machine learning repository (Asuncion & Newman, 2007) to study their inner dependency structure. Both datasets are complete and have continuous and discrete attributes.

5.2.1 Abalone

Abalone dataset was built to predict the age of abalone based on physical measurements of abalone body, such as weights, height. It composes of 4177 samples with 8 attributes. It can be viewed as a regression problem where some measurements are possibly intrinsically interrelated. Instead of predicting the age, we focus on the dependence relations among attributes, which may benefit the prediction task.

There are a few outliers in the dataset. They are usually eliminated by pre-processing step otherwise they may cause large deviation in the following dependence measures calculation. But here it is unnecessary because copula is less susceptible to outliers. This makes

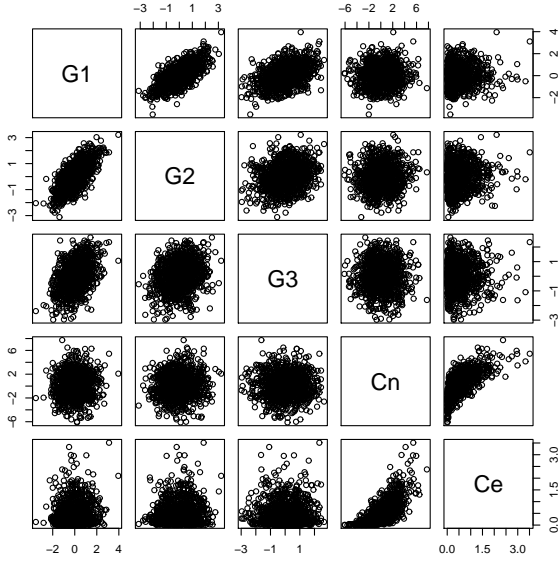


Figure 1: Samples in the simulated data experiment. ‘G1-3’ represent Gaussian, and ‘Cn’, ‘Ce’ represent two copula variable with normal and exponential margins.

copula more advantageous than other moment-based dependence measures sensitive to outliers. When estimating empirical copula in the experiment, we set the order of lattice with different size empirically considering a trade-off between approximation accuracy and computational cost. We choose Spearman’s rho, which is calculated by equation (10), and MI as distance measurements. During MI estimation, kernel method with well-tuned parameters was applied on different moderate sized subsets randomly sampled from the whole dataset. The estimation value varies a little. Due to space limitation, we will give no details about that. Then with these two types of dependence measures as weights, MSP trees were built.

The original dataset are plotted in Figure 3. For illustrative propose, we only present a subset attributes containing four attributes. During empirical copula estimation, the effect of outliers diminishes, which can be easily learned from a comparison between Figure 3 and Figure 4.

Besides robustness to outliers, we emphasis another effect made possible by copula that dependence relations can be successfully revealed by the estimated copula. It can be observed from Figure 3 that all the attributes possesses non-gaussianity to some extent which is demonstrated in their joint densities with other attributes. While all the pairwise estimated cop-

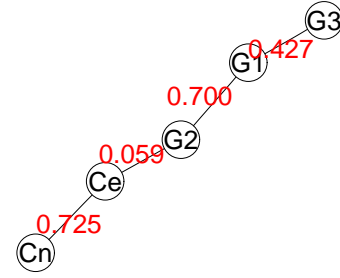


Figure 2: Dependence tree estimated in the simulated experiment.

ulas seems to show a very similar dependency structure after individual properties of variables are believed to be separated away from joint distribution.

Figure 5 shows one of all the maximum spanning trees for abalone in our experiments, where edges are labeled by weights. Except Sex and rings, seven other attributes are linked with relatively strong weighted edges, which is illustrated in Figure 4. It can also be learned that the edges linked the seven physical measurements are the backbone of all the estimated trees, while the nodes for “sex” and “ring” are leaves randomly attached to this seven nodes. This can be interpreted as the reflection of abalone’s body growth. That is, all the physical indexes increase as the abalones grow up, while ring and sex is not strongly related with these seven physical attributes. We argue that predicting ring with the other attributes in abalone dataset may not be a good experimental design.

5.2.2 Housing

The Boston house price dataset is from a 1970 census, first published by Harrison, D. and Rubinfeld (Harrison & Rubinfeld, 1978), with the aim to study how to predict “Medv”¹ based on the 13 attributes. It contains 506 samples, with 14 mixed type attributes, including 13 continuous attributes and 1 binary one. Previous research mainly treat it as a regression problem where the interrelation between attributes are ignored. In our experiment, we studied the whole dependence structure instead. Using copula to estimate dependence relations and to generate a maximum weight tree, we find some unnoticed relations between the at-

¹The abbr. of 14 attributes of Housing dataset refer to UCI machine learning dataset website. (Asuncion & Newman, 2007)

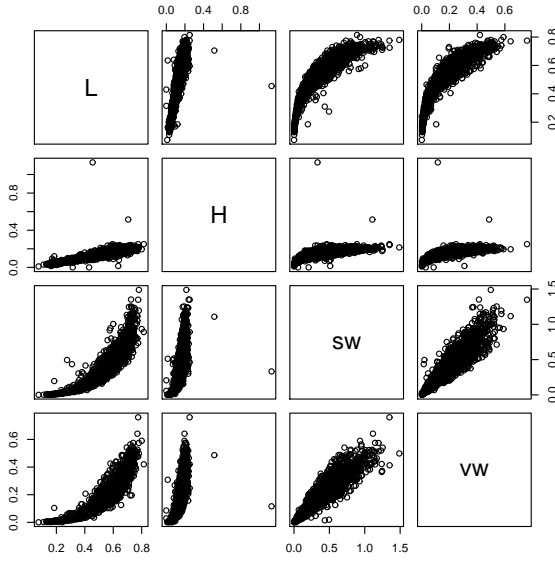


Figure 3: Scatter plot of four attributes in Abalone dataset. ‘L’, ‘H’, ‘sw’, ‘vw’ represent Length, Height, Shucked weight, and Viscera weight.

tributes.

Some researchers propose to transform the data into a suitable scale before further dependence analysis, through monotonically increasing function, such as normalization, nonlinear exponential/log functions. In our experiment, it is unnecessary due to copulas invariant to such kind of transformation. As the previous abalone experiment, MSP algorithm was run on the moderate datasets randomly sampled from housing dataset. Many dependence trees were generated, one of which is plotted in Figure 7. Experimental results indicate that only two link edges including “crim-rad” and “medv-lstat”, remain stable in all the estimated trees. We also observed that there are two group of attributes² are weakly interconnected to some extent.

6 Discussions

Our philosophy on structure learning is that the more we know, the better structure we can learn. Learning by graphical models has its limitations because it is based on only first order dependence relations. This kind of relations are from empirical causality and is suitable for understanding simple mechanism, but probably fail to complicated situations. Using copula, one can incorporate all the dependence information

²One includes “nox,dis,indus,crim”; the other includes “medv,lstat,age,ptratio”.

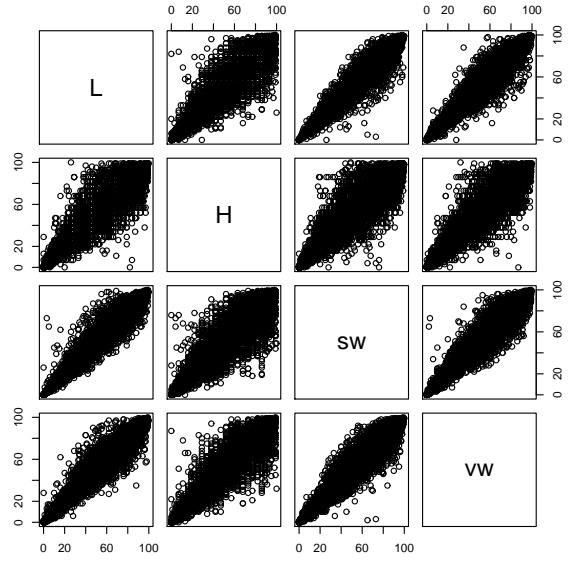


Figure 4: The estimated empirical copula of four attributes in Abalone dataset. For names of variables, see Figure 3

without model constraints and meanwhile all the information is nothing else but dependence relations. Then structure learning based on copula provides a general framework which can unify all the related structure learning methods. The main advantage is its non relevance to particular properties of individual variables.

As we remarked above, different densities may have same copulas. In this paper, the goal of structure learning based on copula is to maximize a total sum of certain dependence measures of structure to span the structure as large as possible. The same task can be achieved by maximum likelihood method, fitting in a sense of mean-squared error. The generalization of tree structure to more complex structure, such as clusters, or hypergraph can be done in a similar way. The type of the structure may be determined on priors and application backgroups.

Dependence representation using only bivariate dependence is limited. Given $N(N - 1)$ pair dependence relations of N random variable, only $N - 1$ of relations compose of tree approximation. To examine the degree of approximation, we propose a criteria by a ratio of total dependence relation of tree ratio to the sum of all the $N(N - 1)$ relations. The ratio of all the experiment are plotted in Figure 8. It can be learned that such tree approximation possesses a large portion of the total weight with relatively few edges meanwhile there is also a large amount of dependence relations

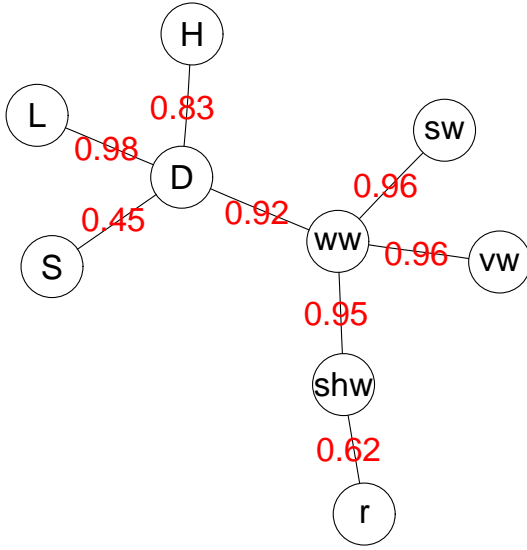


Figure 5: Maximum spanning copula generated from the estimated empirical copula of Abalone dataset.

not included in it. This indicates that through capable of grasp some major first-order dependence/causal relations, bivariate dependence structure itself has a limited capability of dependence modeling on complex cases.

7 Conclusions and Further Directions

In the paper, we propose estimating dependence structure using copula method. Copula can represent all kinds of dependence relations among random variables, and makes no additional assumption on the underlying distributions. Graphical models is a special case in copula family named product copula. A ranked-based algorithm for copula estimation is presented, such that copula function is separated from the joint density with properties of individual variables. Such a nonparametric method can provide very large freedom to structure learning in that the estimated empirical copula contains all the dependence information in the data. Then we study learning product copula or tree structure based on empirical copula. a Chow-Liu like method based on empirical copula is proposed to estimate maximum spanning product copula with only bivariate dependence relations. The proposed method was applied on simulated data and two real dataset to approximate their dependence structure. Experimental results show that copula can achieve a margin-free dependence representation, robust to outlier, and invariant to increasing transforma-

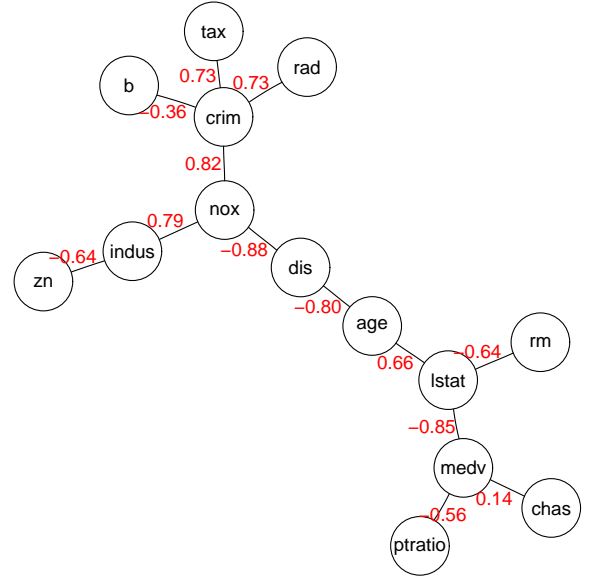


Figure 6: Maximum weight tree with correlation as weight, generated from Housing dataset.

tion and the estimated product copulas can benefit us understanding the underlying dependency structure.

Though widely applied to financial problems in the past few decades, copula theory itself is still probably in its infancy and gradually gaining the attention of many related fields. The most significance of copula to machine learning is that it provides a more general way of dependence representation, measurement, and inference than the previous proposed models. In this paper, we only contribute a little on this issue compared with copula's potential. As far as structure learning is concerned, many problems remains for copula methods. For example, how to choose copula model(or family when in a parametric way) for different applications? How to design copula model for particular dependence relations? Product copula represents the simplest dependence relations. In further work, we suppose to study the estimation of more complex dependence structures through copula. This is very important to many real applications in biology, finance, and social science.

References

- Asuncion, A., & Newman, D. (2007). UCI machine learning repository.
- Buntine, W. (1996). A guide to the literature on learning probabilistic networks from data. *IEEE*

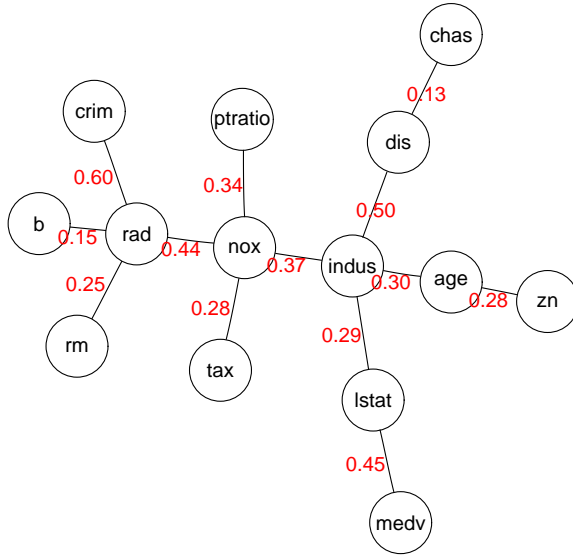


Figure 7: Maximum weight tree with mutual information as weight, generated from Housing dataset.

Transactions on Knowledge and Data Engineering, 8, 195–210.

Chow, C., & Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14, 462–467.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. John Wiley & Sons.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés – un test non paramétrique d’indépendance. *Académie Royale de Belgique - Bulletin de la Classe des Sciences - 5e Série*, 65, 274–292.

Deheuvels, P. (1981). A non parametric test for independence. *Publications de l’Institut de Statistique de l’Université de Paris*, 26, 29–50.

E. Bouyé, V. Durrleman, A. N. G. R., & Roncalli, T. (2000). Copulas for finance - a reading guide and some applications.

Harrison, D., & Rubinfeld, D. (1978). Hedonic prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81–102.

Heckerman, D., Geiger, D., & Chickering, D. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20, 197–243.

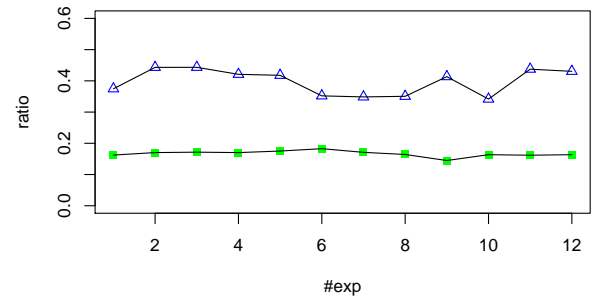


Figure 8: the ratio of total weight of estimated MSP to the sum of all the dependence relations of two experiments on Abalone and Housing. Rectangle represents abalone and triangle represents housing.

Joe, H. (1997). *Multivariate models and dependence concepts*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Jordan, M. (1998). *Learning in Graphical Models*. Kluwer Academic Publishers.

Kirshner, S. (2007). Learning with tree-averaged densities and distributions. *Advances in Neural Information Processing Systems, NIPS 2007*.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7, 48–50.

Ma, J., & Sun, Z. (2007). Copula component analysis. *Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation, ICA2007*.

Nelsen, R. (1998). *An introduction to copulas*. Lecture Notes in Statistics. Springer, New York.

Prim, R. (1957). Shortest connection networks and some generalisations. *Bell System Technical Journal*, 36, 1389–1401.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Sklar, A. (1959). Fonctions de repartition à n dimensions et leurs marges. *Publications de l’Institut de Statistique de l’Université de Paris*, 8, 229–231.

